Expert Systems

Article

DOI: 10.1111/j.1468-0394.2012.00646.x

Effects of the correlation model, the trend model, and the number of training points on the accuracy of Kriging metamodels

Erdem Acar

Department of Mechanical Engineering, TOBB University of Economics and Technology, Söğütözü Cad. No: 43, Ankara, 06560, Turkey
Email: acar@etu.edu.tr

Abstract: *This paper explores the effects of the correlation model, the trend model, and the number of training points on the accuracy of Kriging metamodels. Gaussian correlation models are found to be superior to exponential and linear correlation models. No particular trend model is found to be better than the other models. The number of training points used in constructing the Kriging metamodels is observed to change the relative performances of the trend and the correlation functions. The leave-one-out cross-validation error is found to become a better surrogate for the actual error, as the number of training points is increased. Finally, the use of an ensemble of metamodels is discussed and it is found that using an ensemble may improve the accuracy.*

Keyword: accuracy, Kriging metamodel, ensemble

1. Introduction

Computer simulations are used in many disciplines including astrophysics, biology, chemistry, economics, engineering, entertainment, and social sciences. For instance, in the engineering discipline, design of advanced vehicles (e.g. aircraft, automobiles) relies on high-fidelity computer simulations for accurate analysis of system characteristics. Simulation models of acceptable accuracy have required at least 6–8 h of CPU time throughout the last 30 years, even though computer processing power along with memory and storage capacities have drastically increased (Venkataraman & Haftka, 2004). This lack of computation speed-up can be explained by the fact that the fidelity and the complexity of the models have also steadily increased over the same period. When these computationally expensive high-fidelity simulations are used within a design optimization framework, the computational cost becomes excessive. In addition, if gradient-based techniques are used for optimization, the accuracy and the convergence of the optimization solution can suffer, since the objective function or the constraint functions can be noisy. Hence, it is common practice to replace the computationally expensive simulations with smooth analytic functions that can serve as surrogate models for efficient response estimation. These approximate models are also called metamodels, emulators, etc.

Metamodelling techniques aim at approximating the response data at the specified training points that are selected using a design of experiments technique (e.g. central composite design, Latin hypercube design). There are many metamodelling techniques including polynomial response surface approximations (Myers *et al*., 2009), multivariate adaptive regression splines (Friedman, 1991), radial basis functions (Hardy, 1971; Mullur & Messac, 2005), neural networks (Smith, 1993), support vector regression (Gunn, 1997; Clarke *et al*., 2005), and Kriging (Sacks *et al*., 1989; MacKay, 1998; Simpson *et al*., 2001b; Lophaven *et al*., 2002; Martin & Simpson, 2005; Wang *et al*., 2005).

This study focuses on Kriging metamodels that have shown successful applications in many disciplines of engineering design. In aerospace design, amongst many examples, Kriging is used for optimization of an aerospike nozzle problem (Simpson *et al*., 2001a), aerodynamic design of centrifugal compressor's impeller (Wang *et al*., 2006), and performing multi-objective design exploration for a three-element airfoil consisted of a slat, a main wing, and a flap (Kanazaki *et al*., 2007). The Kriging applications in other engineering disciplines include the studies for electromagnetic device optimization (Lebensztajn *et al*., 2004), automobile crashworthiness design (Forsberg & Nilsson, 2005), structural reliability analysis (Kaymaz, 2005), structural analysis of concrete dams (McLean *et al*., 2006), and structural optimization of an automotive door (Lee & Kang, 2007).

The trend and the correlation functions used in Kriging metamodels affect the accuracy of the constructed metamodels. There is limited guidance in the literature on selecting the form of the trend and the correlation functions (Martin & Simpson, 2005). Stein (1999) noted that the Matern family of functions is a good candidate. Simpson *et al*. (2001b) noted that zeroth-order trend and Gaussian correlation functions are the most commonly used models in engineering applications. The main contribution of this paper is to address the issue of how to improve the accuracy of Kriging metamodels. For this purpose, the effects of choosing the proper trend and correlation models, the effects of the number of training points, the effectiveness of cross-validation error to represent the actual error at the test points, and the use of an ensemble of Kriging metamodels are investigated.

The remainder of the paper is organized as follows. Section 2 presents the basics of Kriging metamodels. Section 3 first discusses different trend and correlation models used in Kriging metamodels. Then, a weighted sum formulation is presented in Section 3 that combines various Kriging metamodels with different combinations of trend and correlation models in the form of an ensemble. Section 4 describes the benchmark problems considered in this study. The results of this study are presented and discussed in Section 5, followed by some concluding remarks provided in Section 6.

2. Kriging metamodels

The basic assumption in Kriging metamodels is that the estimation of a response function y, expressed in terms of the input variables x, is in the following form:

$$\hat{y}(x) = \mathbf{p}^{\mathrm{T}}(x)\boldsymbol{\beta} + Z(x) \quad (1)$$

The trend model $\mathbf{p}^{\mathrm{T}}(x)\boldsymbol{\beta}$ is a polynomial of given order (e.g. first order in Figure 1) that globally approximates the response. The departure model $Z(x)$ is the stochastic component that generates deviations such that the Kriging model interpolates the sampled response data (see Figure 1).

Amongst the trend models, the zeroth-order (i.e. a constant value) (Simpson *et al.*, 2001b) and the first-order (i.e. linear) (Zerpa *et al.*, 2005) polynomials are the most commonly used models. Figure 2 depicts the Kriging metamodels with constant and linear trend models over the same data points. In this study, the use of a second-order (i.e. quadratic) polynomial is also considered.

The stochastic component has a mean value of zero and the following covariance:

$$COV\left[Z\left(\mathbf{x}^i\right), Z\left(\mathbf{x}^j\right)\right] = \sigma^2\mathbf{R}\left[R\left(\mathbf{x}^i, \mathbf{x}^j\right)\right] \quad (2)$$

where $\mathbf{R}$ is an $N \times N$ correlation matrix if N is the number of data points, $R\left(\mathbf{x}^i, \mathbf{x}^j\right)$ is the correlation function between the two data points x^i and x^j. The correlation models considered in this study have the following mathematical form:

$$R(\theta, d) = \prod_{k=1}^{L} R_k\left(\theta_k, d_k\right) \quad (3)$$

where R_k is the correlation function defined in terms of the unknown model parameters θ_k, and the distance d_k between the kth components of the two data points $\mathbf{x}^i$ and $\mathbf{x}^j$ ($d_k = x_k^i - x_k^j$), and L is the number of input variables. The most commonly used correlation function is the Gaussian function (Wang *et al.*, 2005), which has the form of

$$R(\theta) = \prod_{k=1}^{L} \exp\left(-\theta_k d_k^2\right) \quad (4)$$

In this paper, two additional correlation models (exponential and linear) are also considered. The mathematical forms of the exponential and the linear correlation models are given in equations (5) and (6), respectively.

$$R(\theta) = \prod_{k=1}^{L} \exp\left(-\theta_k \left|d_k\right|\right)(\text{exponential}) \quad (5)$$

$$R(\theta) = \prod_{k=1}^{L} \max\left\{0, 1 - \theta_k \left|d_k\right|\right\}(\text{linear}) \quad (6)$$

The variation of the correlation function, R_k, with the distance, d_k, for exponential, Gaussian, and linear functions is depicted in Figure 3.

Once the correlation function has been estimated, the response y in equation 1 is predicted as

$$\hat{y}(x) = \mathbf{p}^{\mathrm{T}}(x)\hat{\beta} + \hat{\mathbf{r}}^{\mathrm{T}}(x)\hat{\mathbf{R}}^{-1}\left(\mathbf{F} - \mathbf{P}\hat{\beta}\right) \quad (7)$$

where the vectors $\hat{\mathbf{r}}$ and $\hat{\beta}$ are given by

$$\hat{\mathbf{r}}^{\mathrm{T}}(x) = \left[\hat{R}\left(x, x^1\right), \hat{R}\left(x, x^2\right), \ldots, \hat{R}\left(x, x^N\right)\right]^T \quad (8.1)$$

$$\hat{\beta} = (\mathbf{P}^{\mathrm{T}}\hat{\mathbf{R}}^{-1}\mathbf{P})^{-1}\mathbf{P}^{\mathrm{T}}\hat{\mathbf{R}}^{-1}\mathbf{F} \quad (8.2)$$

where $\hat{\mathbf{r}}^{\mathrm{T}}(x)$ is the correlation vector of length N between a prediction point x and the N sampling points, the $\hat{\mathbf{R}}$ matrix is obtained by using the predicted values $\hat{\theta}_k$ in equation (3), $\mathbf{F}$ represents the responses at the N points, and $\mathbf{P}$ is obtained by evaluating $\mathbf{p(x)}$ array at the N points.

The variance of the output y (or the variance of Z in equation (1) can be estimated as

$$\hat{\sigma}^2 = \frac{\left(\mathbf{F} - \mathbf{P}\hat{\beta}\right)^T\hat{\mathbf{R}}^{-1}\left(\mathbf{F} - \mathbf{P}\hat{\beta}\right)}{N} \quad (9)$$

The estimator of the predictor variance is presented in literature (see Den Hertog *et al.*, 2006), but this paper does not use the predictor variance; instead it uses the cross-validation error.

The unknown model parameters θ_k can be estimated by solving the following constrained maximization problem (Lophaven *et al.*, 2002):

$$\begin{aligned} &Max \;\; \Phi(\Theta) = -\left[N\ln(\hat{\sigma}^2) + \ln|\mathbf{R}|\right] \\ &s.t. \quad \Theta > 0 \end{aligned} \quad (10)$$

where $|\mathbf{R}|$ is the determinant of $\mathbf{R}$, Θ is the vector of unknown parameters θ_k, and both $\hat{\sigma}$ and $\mathbf{R}$ are functions of Θ. In this study, the MATLAB Kriging toolbox developed by Lophaven *et al.* (Martin & Simpson, 2005) is used.

This paper focuses on single output models. When a multiple output system is of interest, each output can be treated independently while ignoring the correlation between them. In literature, however, there exist some studies that account

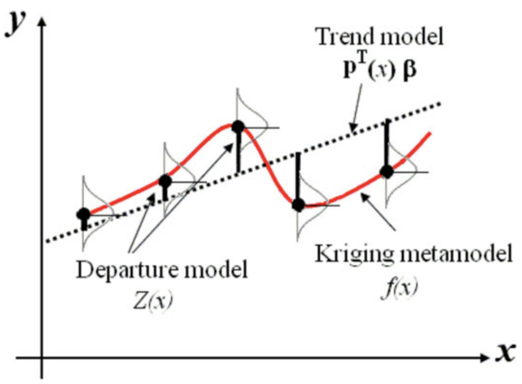

Figure 1: *Response approximation via Kriging metamodel*

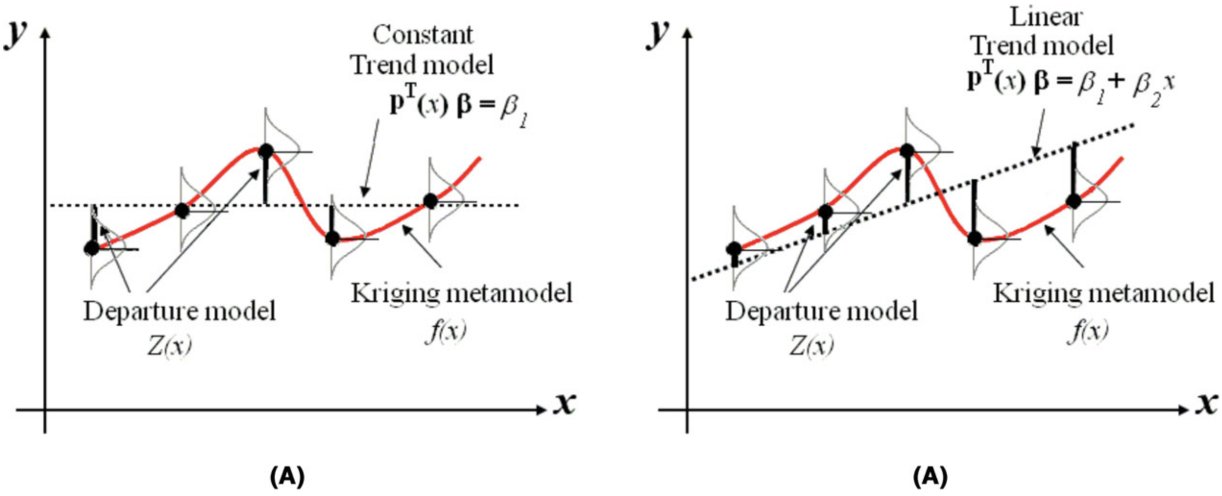

Figure 2: *Kriging metamodels with (A) constant and (B) linear trend model.*

for the correlation between the outputs (Santner *et al.*, 2003; Li *et al.*, 2006).

3. An ensemble of different Kriging models

The predictive capability of Kriging metamodels may depend on the trend model and the correlation model used. The traditional approach is to select a trend model and a correlation model based on experience or intuition. A more judicious way to choose trend and correlation models may be as follows. First, various Kriging metamodels with different combinations of trend models (e.g. constant, linear, quadratic) and correlation functions (e.g. exponential, Gaussian, linear) are constructed. Then, the combination with the greatest accuracy is usually chosen, while the other combinations are discarded. Instead, all the constructed Kriging metamodels can be combined in the form of an ensemble, which takes advantage of the prediction ability of each individual metamodel to increase the accuracy of the predicted response. The reader is referred to other studies (Zerpa *et al.*, 2005; Goel *et al.*, 2007; Acar & Rais-Rohani, 2009) for more information on the ensemble of metamodels. The remainder of this section provides brief information on the ensemble of Kriging metamodels.

The idea of using an ensemble of Kriging metamodels is borrowed from the work of Bishop (1995) on neural networks. Bishop (1995) combined different neural networks in the form of a committee to take advantage of the predictive capabilities of the individual neural networks. In this work, Kriging metamodels composed of different combinations of trend and correlation models are combined in the form of an ensemble using a weighted sum formulation:

$$\hat{y}_{ens}(x) = \sum_{i=1}^{M} w_i(x)\,\hat{y}_i(x) \tag{11}$$

where $\hat{y}_{ens}$ is the ensemble prediction for the response, M is the number of individual Kriging metamodels used in the ensemble, w_i is the weight factor for the ith individual Kriging metamodel ($0 \le w_i \le 1$), $\hat{y}_i$ is the response estimated by the ith Kriging metamodel, and x is the vector of independent input variables.

The weight factors in equation (11) are calculated while satisfying the requirement

$$\sum_{i=1}^{M} w_i(x) = 1 \tag{12}$$

In general, the weight factors are selected such that the metamodels with high accuracy have large weight factors and vice versa. Bishop (1995) proposed selecting the weight factors as

$$w_i = \sum_{j=1}^{M} \left(\mathbf{C}^{-1}\right)_{ij} \Bigg/ \sum_{m=1}^{M}\sum_{j=1}^{M} \left(\mathbf{C}^{-1}\right)_{mj} \tag{13}$$

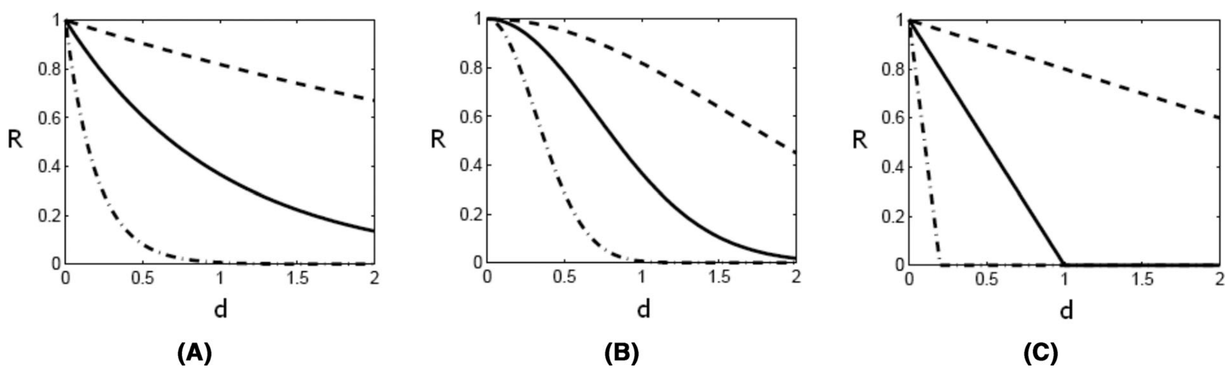

Figure 3: *Correlation functions for Kriging metamodels: (A) exponential, (B) Gaussian, and (C) linear function. Here, the values of θ_k are 0.2, 1.0, and 5.0 for the dashed line, the solid line, and the dash-dotted line, respectively.*

where **C** is the estimated error covariance matrix whose elements are calculated from

$$C_{ij} = \frac{1}{N}\sum_{k=1}^{N}\left(\hat{y}_i^k - y^k\right)\left(\hat{y}_j^k - y^k\right) \tag{14}$$

where y^k is the true response value corresponding to the input vector x_k, with $\hat{y}_i^k$ and $\hat{y}_j^k$ the corresponding predicted values by the ith and jth neural networks, respectively.

Selecting weight factors from equation (13) minimizes the error in the whole domain of input variables based on the assumption that the errors of different neural networks are uncorrelated and unbiased (that is, with zero mean), which is not always true. In addition, even though this approach may be suitable for application to a committee of neural networks, it may be unsuitable for ensembles based on other metamodelling techniques. In the case of neural networks, since there is a difference between the true response computed at each training point and the prediction of a neural network, $\hat{y}_i^k$, the error correlation matrix **C** is non-vanishing. For Kriging metamodels, however, if the error metric is chosen as the difference between the predicted and the true responses at the training points (see equation (14)), then the difference is zero (because Kriging metamodels are exact interpolators) and the correlation matrix **C** becomes a null matrix. A possible solution to this problem is to use leave-one-out cross-validation errors.

In this study, the weight factors are selected by solving an optimization problem of the form (Acar & Rais-Rohani, 2009)

$$\text{Find } w_i,\ i = 1, \ldots, M \tag{15.1}$$

$$\min\ GMSE = \frac{1}{N}\sum_{k=1}^{N}\left(y^k - \hat{y}_{ens}^{(k)}\right)^2 \tag{15.2}$$

$$\text{s.t.} \sum_{i=1}^{M} w_i = 1 \tag{15.3}$$

where y^k is the true response at x_k and $\hat{y}_{ens}^{(k)}$ is the corresponding predicted value from the ensemble model constructed using all except the kth design point. As evident by (15.2), the greater the number of training points the higher the cost of calculating the GMSE metric. This metric gives an average error in the estimated response at the selected training points. Therefore, depending on the number and the distribution of training points, GMSE may not necessarily provide the evidence of the global error in the whole domain of input variables. One of the objectives of this paper is to investigate the effect of the number of training points on the efficiency of GMSE in providing a global error measure.

4. Test problems and numerical procedure

The following six benchmark functions are used as test problems in this study.

i. *A cubical function* (Mullur & Messac, 2005)

$$y\left(x_1, x_2\right) = 0.5x_1^3 + x_2^2 - x_1x_2 - 7x_1 - 7x_2 \tag{16}$$

where $x_1, x_2 \in [5,\ 10]$.

Table 1: *Parameters used in Hartman-3 function*

i	a_{ij}			c_i	p_{ij}		
1	3.0	10.0	30.0	1.0	0.3689	0.1170	0.2673
2	0.1	10.0	35.0	1.2	0.4699	0.4387	0.7470
3	3.0	10.0	30.0	3.0	0.1091	0.8732	0.5547
4	0.1	10.0	35.0	3.2	0.03815	0.5743	0.8828

ii. *Branin-Hoo* (Dixon & Szego, 1978)

$$y\left(x_1, x_2\right) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right) \times \cos\left(x_1\right) + 10 \tag{17}$$

where $x_1 \in [-5,\ 10]$, and $x_2 \in [0,\ 15]$.

iii. *Camelback* (Dixon & Szego, 1978)

$$y(x_1, x_2) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + \left(-4 + 4x_2^2\right)x_2^2 \tag{18}$$

where $x_1 \in [-3,\ 3]$, and $x_2 \in [-2,\ 2]$.

iv. *Goldstein-Price* (Dixon & Szego, 1978)

$$y\left(x_1, x_2\right) = \left[1 + \left(x_1 + x_2 + 1\right)^2\left(19 - 4x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2\right)\right]\left[30 + \left(2x_1 - 3x_2\right)^2\left(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2\right)\right] \tag{19}$$

where $x_1, x_2 \in [-2,\ 2]$.

v. *A sinusoidal function* (Mullur & Messac, 2005)

$$y\left(\mathbf{x}\right) = x_1 \sin\left(x_2\right) + x_2 \sin\left(x_1\right) \tag{20}$$

where $x_1, x_2 \in [-2\pi,\ 2\pi]$.

vi. *Hartman-3* (Dixon & Szego, 1978)

$$y\left(\mathbf{x}\right) = -\sum_{i=1}^{m} c_i \exp\left[-\sum_{j=1}^{n} a_{ij}(x_j - p_{ij})^2\right] \tag{21}$$

where $x_i \in [0,\ 1]$. In this study, the three-variable ($n = 3$) model of this function is considered. The values of function parameters c_i, a_{ij}, and p_{ij} for Hartman-3 function are given in Table 1. The value of the parameter m for both cases is taken as 4.

4.1. Numerical procedure

The first step of constructing a metamodel is to choose the Kriging metamodel parameters, order of polynomial, and type of correlation function. Then, a design of experiments type is selected to determine the points in the design space where the numerical experiments are conducted. The selected points are called training points (or sampling points). Next, numerical experiments are performed at the training points. Then, the Kriging metamodel is constructed using the design of experiments information and the results of numerical experiments, and the accuracy of the metamodel is evaluated using the normalized root mean square error (NRMSE).

In this study, the maximin space-filling technique proposed in Mourelatos *et al.* (2006) is used to generate the training points. The number of training points is varied and its effect on the results is explored. The benchmark functions are evaluated at the selected training points, and Kriging metamodels with various trend and correlation function combinations (nine combinations overall) are constructed. Then, the accuracies of the constructed metamodels are evaluated using a large number of test points generated on a uniform grid. For the two-dimensional functions, $32^2 = 1024$ test points, and for the three-dimensional function $10^3 = 1000$ test points are used. NRMSE is used as the error metric to quantify the accuracy of the metamodels

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{N_t} \left[(y_t)_i - (\hat{y}_t)_i\right]^2}{\sum_{i=1}^{N_t} (y_t)_i^2}} \tag{22}$$

where N_t is the number of test points, $(y_t)_i$ and $(\hat{y}_t)_i$ are the true function value and the metamodel prediction of the function value at the ith test point, respectively. The difference between equations (15.2) and (22) is that the cross-validation error is used in equation (15.2) while the error at test points is used in equation (22).

It should be noted that the use of 1000 (approximately) test points is only feasible for these academic mathematical test problems. For real-life problems, since the evaluation of a response function value takes about 6–8 h (Venkataraman & Haftka, 2004), using 1000 (or 1024) test points is not feasible. Instead, either a small number of test points is used (which can only give a questionable accuracy) or no test points are used at all but cross-validation error is used. In this study, normalized cross-validation error, denoted as $NRMSE_{CV}$, is also used, and the efficiency in representing the actual error is investigated

$$NRMSE_{CV} = \sqrt{\frac{\sum_{i=1}^{N} \left[y_i - \hat{y}_i^{-i}\right]^2}{\sum_{i=1}^{N} (y_i)^2}} \tag{26}$$

where $\hat{y}_i^{-i}$ is the function prediction of the Kriging metamodel constructed without using the ith training point.

5. Results

In this section, first the effects of the trend and the correlation models on the accuracy of the Kriging metamodels are investigated. Next, the effect of the number of training points on the relative performances of the trend and the correlation models is analysed. Then, the inter-relation of the cross-validation error and the actual error is explored. Finally, the accuracy of the best individual Kriging metamodel is compared to that of the ensemble of Kriging metamodels.

5.1. Effect of the trend model

The effects of the trend model on the accuracy of Kriging metamodels for the Branin-Hoo and the Camelback functions (chosen as representative examples) are depicted in Figures 4–6. The exponential, the Gaussian, and the linear correlation models are used, respectively.

When the exponential correlation function is used, Figure 4 shows that the zeroth-order trend model worked best for the Branin-Hoo function, while the quadratic model worked best for the Camelback function.

When the Gaussian correlation function is used, Figure 5 shows that the quadratic trend model worked best for the Branin-Hoo function, while the constant and the linear models both worked best for the Camelback function.

When the linear correlation function is used, Figure 6 shows that the quadratic trend model showed a clear superiority over the constant and the linear trend models for both the Branin-Hoo and the Camelback functions.

Similarly, for the other benchmark functions, it is found that the performance of the trend models depends on the correlation function used. No particular trend model is found to outperform the other models for all test problems. It is also found that the number of training points influences the effectiveness of the trend models.

5.2. Effect of the correlation model

The effects of the correlation model on the accuracy of Kriging metamodels for the Branin-Hoo and the Camelback functions (chosen as representative examples) are depicted in Figures 7–9. The constant, the linear, and the quadratic trend models are used, respectively.

Figures 7–9 show for the Branin-Hoo and the Camelback functions that the Gaussian correlation model outperforms the other correlation models, and the linear correlation model displays the worst performance. These findings are also valid for all the test functions investigated. Hence, regardless of the trend model chosen, the Gaussian correlation model performs best, and the linear correlation model performs worst.

5.3. Effect of the number of training points

In general, as the number of training points increased, the accuracy of the metamodels improves. In addition, the number of training points may also affect the relative performances of the trend and the correlation models. The effect of the number of training points on the relative performances of the trend models for the Goldstein–Price function is presented in Figure 10(A). When the number of training points is smaller than 35, the quadratic trend model outperforms all other trend models. As the number of training points is larger than 35, on the other hand, the constant and the linear models outperform the quadratic trend model. Similarly, Figure 10(B) shows for the Hartman-3 function that the linear and the quadratic trend models perform better than the constant model when the number of training points is smaller than 35, while the constant model performs best when the number of training points is larger than 35.

Figure 11(A) depicts the effect of the number of training points on the relative performances of the correlation models for the Goldstein-Price function. The Gaussian correlation model performs best when the number of training points is between 10 and 20, the exponential correlation model performs best when the number of training points is between 20 and 35, and again the Gaussian correlation model

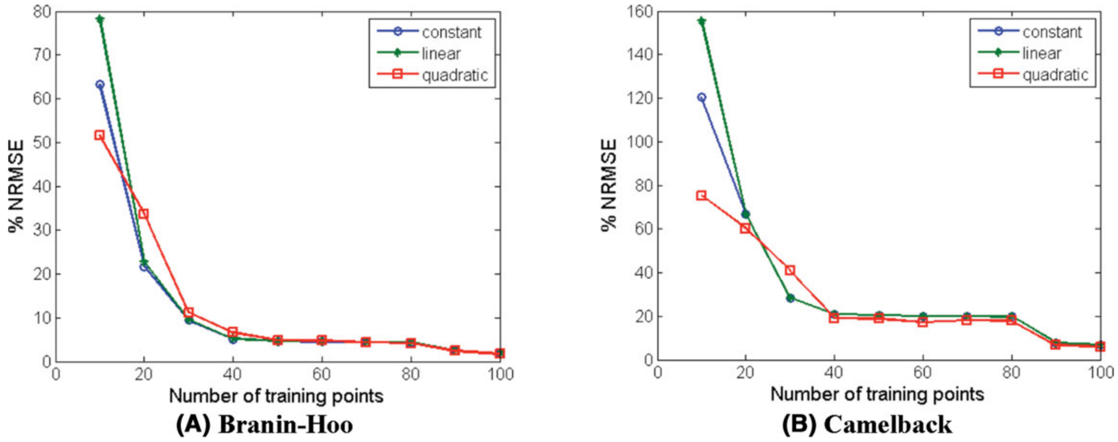

Figure 4: *Effects of the chosen trend model on the accuracy of the constructed Kriging metamodels when exponential correlation model is used in the metamodels.*

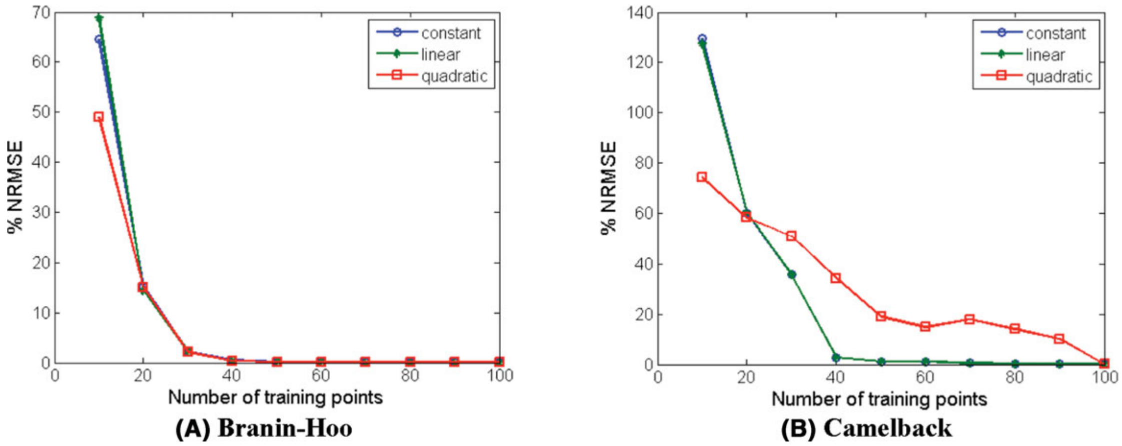

Figure 5: *Effects of the chosen trend model on the accuracy of the constructed Kriging metamodels when Gaussian correlation model is used in the metamodels.*

performs best when the number of training points is larger than 35. Similarly, for the Hartman-3 function, the Gaussian correlation model performs best when the number of training points is between 20 and 45, the exponential correlation model performs best when the number of training points is between 45 and 60, and the Gaussian correlation model performs best when the number of training points is larger than 60.

5.4. Inter-relation of the cross-validation error and the actual error

Inter-relation between the cross-validation error and the actual error for the Branin-Hoo function (chosen as the representative example) are displayed in Figure 12. Since the linear correlation model is found to be the worst choice, it is excluded from the analysis. Overall, three trend models and two correlation models (six combinations) are considered,

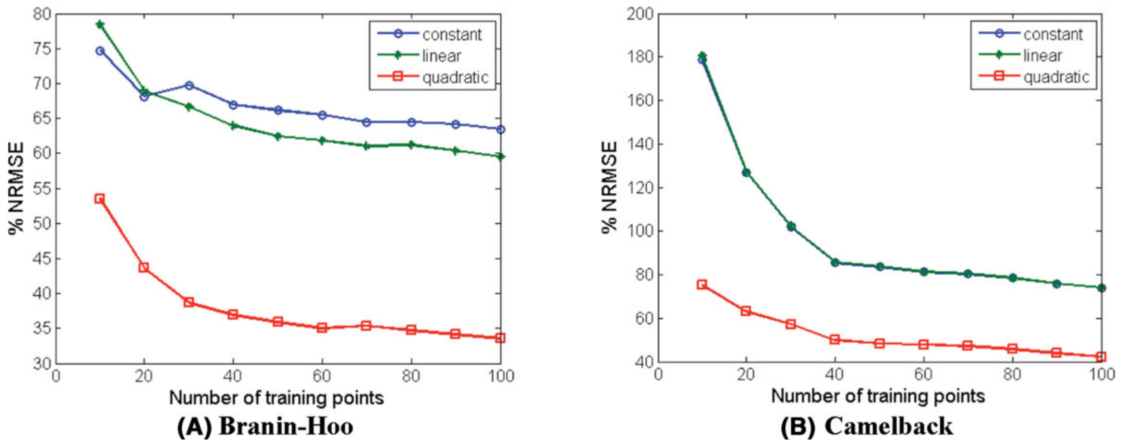

Figure 6: *Effects of the chosen trend model on the accuracy of the constructed Kriging metamodels when linear correlation model is used in the metamodels.*

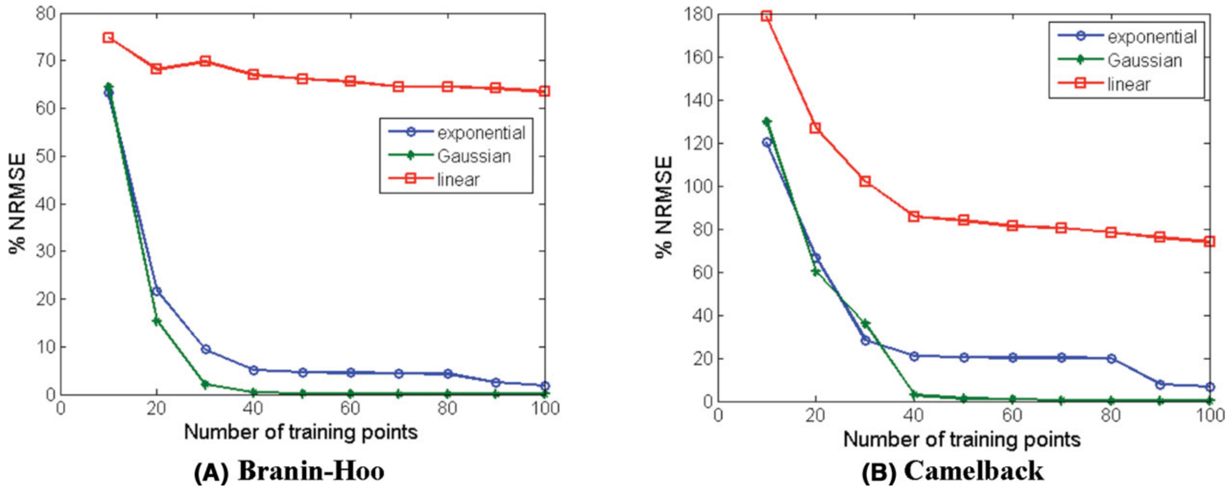

Figure 7: *Effects of the chosen correlation model on the accuracy of the constructed Kriging metamodels when zeroth-order trend model is used in the metamodels.*

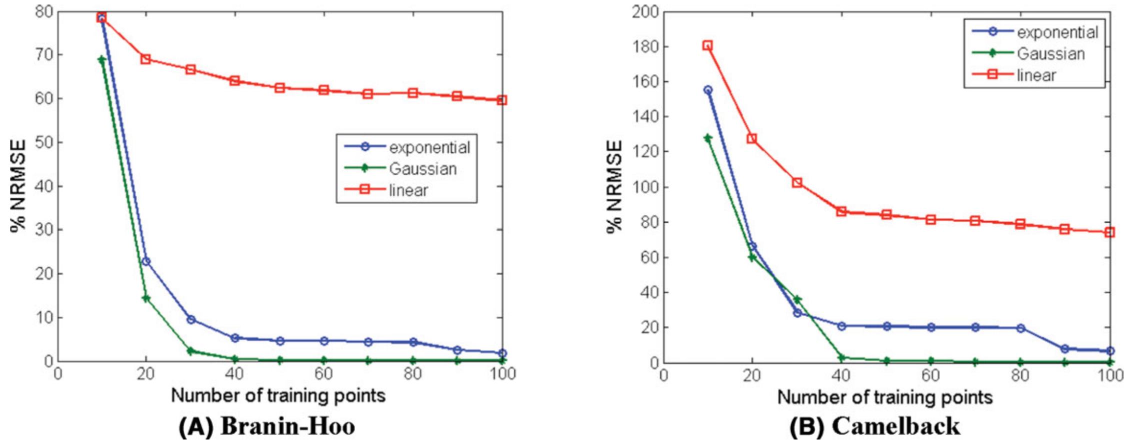

Figure 8: *Effects of the chosen correlation model on the accuracy of the constructed Kriging metamodels when linear trend model is used in the metamodels.*

shown in Figures 12(A) through (E). It is seen that the general trend of the cross-validation error and the actual error is similar. This finding is also valid for the other benchmark problems. The next section will analyse whether the use of cross-validation or actual errors leads to the selection of the same trend/correlation model combinations.

5.5. The advantages of using an ensemble of Kriging metamodels

Sections 5.1 through 5.3 showed that the performances of the constructed metamodels depend on the selected trend and correlation models as well as the training data set. In general, the most accurate trend and correlation

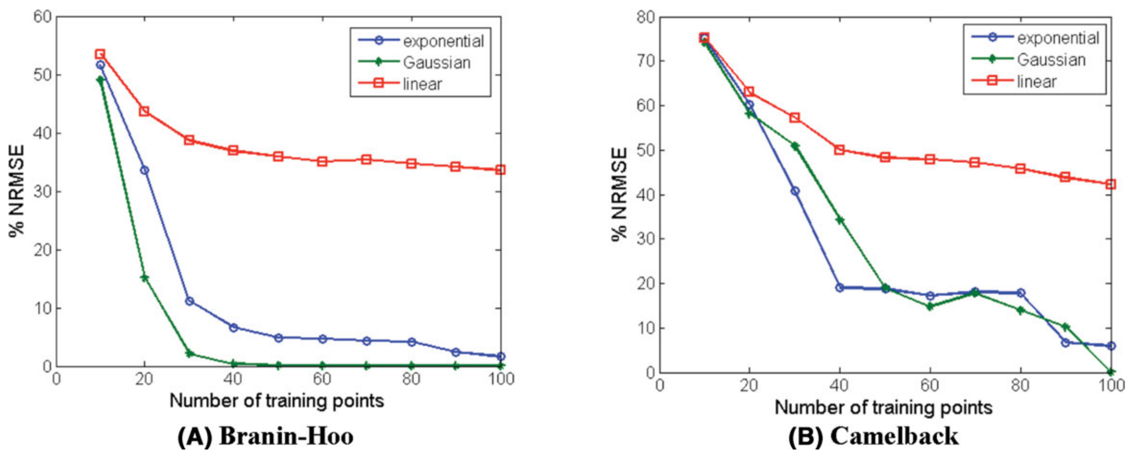

Figure 9: *Effects of the chosen correlation model on the accuracy of the constructed Kriging metamodels when quadratic trend model is used in the metamodels.*

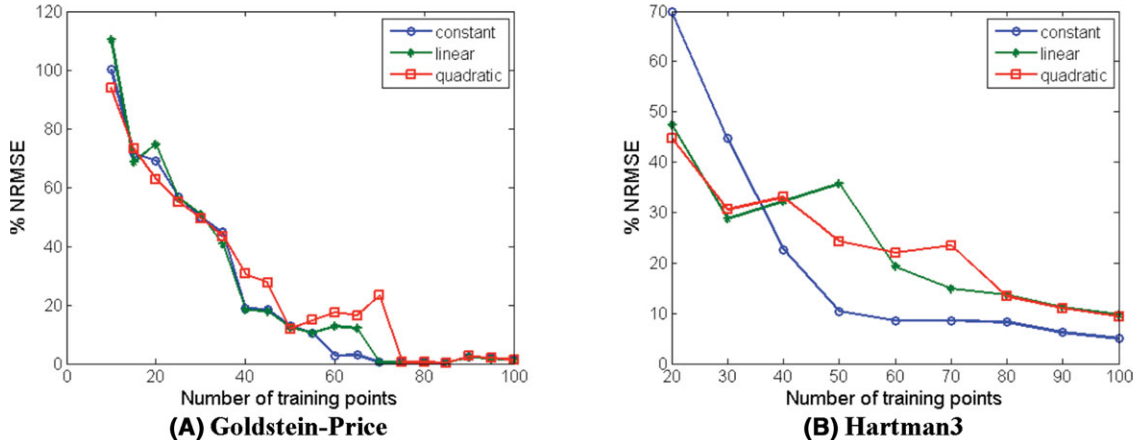

Figure 10: *Effect of the number of training points on the performance of the trend models. Gaussian correlation model is used in the metamodels.*

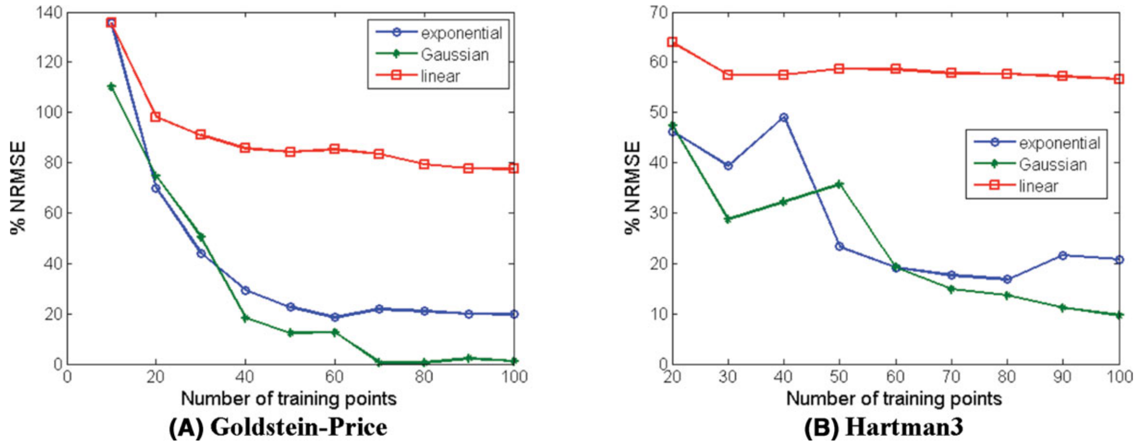

Figure 11: *Effect of the number of training points on the performance of the correlation models. Linear trend model is used in the metamodels.*

models are not known *a priori* for a given problem. Therefore, instead of using a Kriging metamodel with a particular trend and correlation model, it may be more advantageous to use an ensemble of Kriging metamodels. This section analyses the effectiveness of using an ensemble of Kriging metamodels.

For the Branin-Hoo function (chosen as the representative example), the normalized cross-validation errors ($NRMSE_{CV}$) of the individual Kriging metamodels, and the ensemble of metamodels are presented in Table 2. The cross-validation error of the ensemble of Kriging metamodels is smaller than the best individual Kriging metamodel. This result is not very surprising because the ensemble is formed via solving an optimization problem that minimizes the cross-validation error. Therefore, the cross-validation error of the ensemble is always smaller than or equal to that of the best individual metamodel. This finding is also valid for all the benchmark problems investigated.

Instead of cross-validation error comparison, it is more important to compare the actual errors of the metamodels. For the Branin-Hoo function, the actual errors (NRMSE) of the individual Kriging metamodels and the ensemble of metamodels are presented in Table 3. If the number of training points is larger than or equal to 30, the actual error of the ensemble of Kriging metamodels is close to the actual error of the best individual Kriging metamodel.

Comparing Tables 2 and 3 it is seen that if the number of training points are small, different trend/correlation model combinations can be selected as the most accurate whether the cross-validation or the actual error is used. As the number of training points increases, on the other hand, usually the same trend/correlation model combination is selected via the cross-validation or the actual error. This may indicate that the cross-validation error becomes a good surrogate for the actual error as the number of training points increases.

Finally, Table 4 shows the comparison of the accuracy of the ensemble (using NRMSE) to those of the first, the second, and the third most accurate individual metamodels. It is seen that if the number of training points is not sufficiently large, the ensemble is not better than the most accurate metamodel, almost as good as the second most accurate, and always better than the third most accurate metamodel. However, as the number of training points increases, the ensemble becomes as good as the most accurate metamodel.

6. Concluding remarks

This paper analysed the effectiveness of the type of the chosen trend and correlation models, the number of training points, and the use of an ensemble of Kriging metamodels for improving the accuracy of Kriging metamodels. The

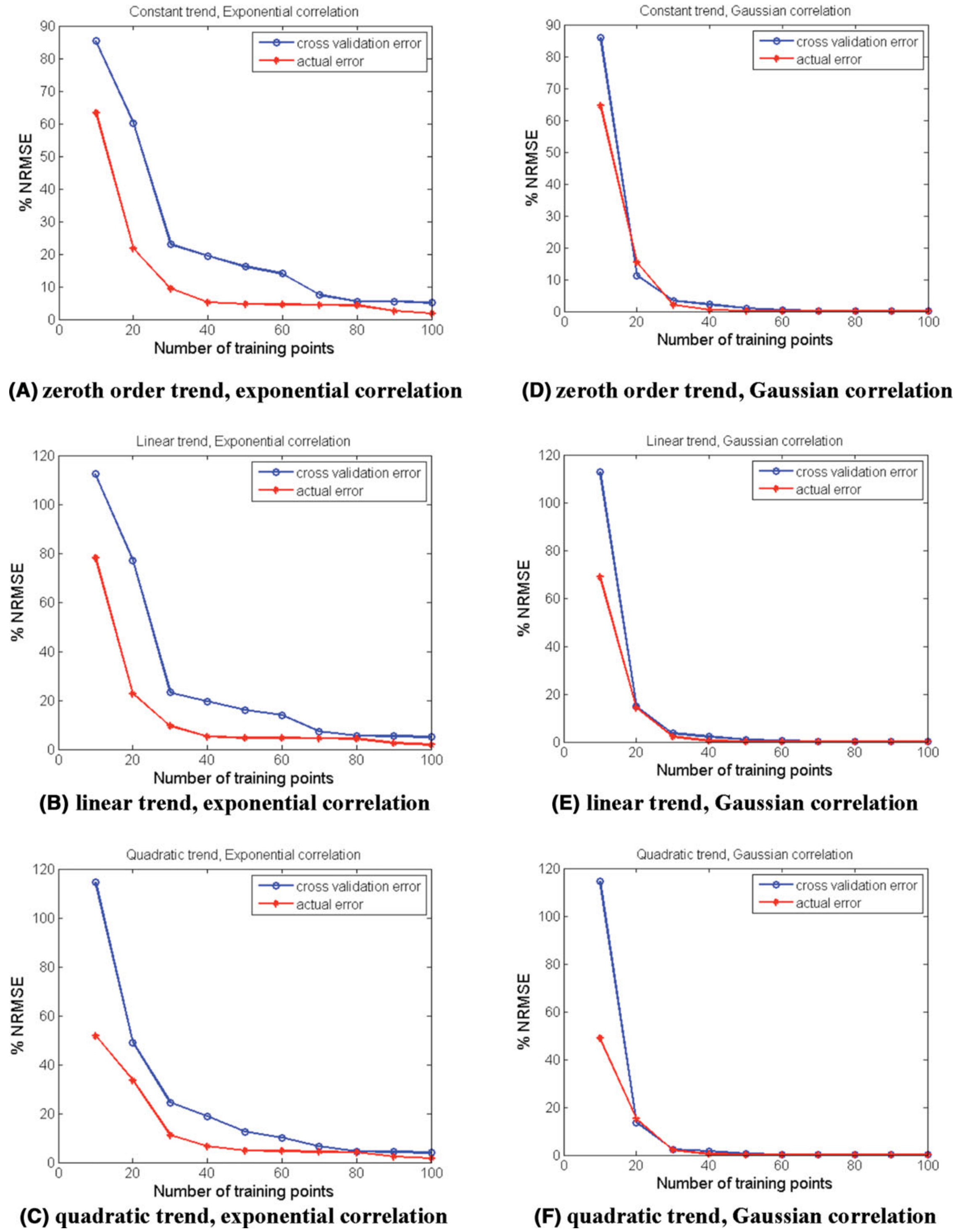

Figure 12: *The variation of the cross-validation error and the actual error for the Branin-Hoo function. Linear correlation models are excluded in comparison.*

investigation on six different mathematical benchmark problems resulted in the following findings.

- For the benchmark problems considered, the Gaussian correlation models resulted in the most accurate results. For the trend models, on the other hand, no particular trend model was found to be superior to the other models.
- The number of training points was found to have a substantial effect on the relative performances of the trend and the correlation functions to one another.
- The effectiveness of cross-validation errors in providing a global error measure was analysed. It was found that as the number of training points is increased, the cross-validation error might become a better surrogate for the actual error.
- The potential of accuracy improvement by using an ensemble of Kriging metamodels is investigated. It was found that the cross-validation error of the ensemble was always smaller than that of the individual metamodel with the smallest cross-validation error. It was also found that that if the number of training points is not sufficiently large, the ensemble is not better than the most accurate metamodel, almost as good as the second most accurate, and always better than the third most accurate metamodel. However,

Table 2: *Normalized cross-validation errors ($NRMSE_{CV}$) for individual Kriging metamodels and the ensemble of metamodels for the Branin-Hoo function. The individual Kriging metamodel with the smallest $NRMSE_{CV}$ is indicated with a bold font*

*N**	*C-E**	*L-E**	*Q-E**	*C-G**	*L-G**	*Q-G**	*C-L**	*L-L**	*Q-L**	*ENS**
10	85.33	112.41	114.66	86.04	112.79	114.56	**82.48**	112.40	114.67	78.49
15	77.24	98.71	85.62	**44.22**	50.90	47.37	80.00	102.69	93.64	40.97
20	60.17	77.09	49.05	**11.29**	14.77	13.56	77.79	91.43	73.95	11.18
25	24.01	24.32	24.28	3.43	**3.14**	3.22	74.69	82.42	58.11	3.06
30	22.97	23.17	24.55	3.34	3.52	**2.45**	72.38	78.53	55.06	2.41
35	20.74	20.99	20.12	2.80	2.93	**1.95**	73.69	79.25	53.02	1.95
40	19.41	19.50	18.84	2.13	2.19	**1.73**	73.95	78.34	52.39	1.65
45	16.71	16.65	13.56	1.34	1.32	**0.68**	72.46	75.89	48.55	0.68
50	16.16	16.09	12.49	0.98	0.92	**0.66**	73.39	75.06	47.91	0.66
55	15.21	15.12	11.75	0.92	0.86	**0.56**	72.61	73.71	48.46	0.56
60	13.96	13.86	10.16	0.46	0.42	**0.19**	72.91	73.21	46.95	0.19

**N*, number of training points; C-E, zeroth-order trend model, exponential correlation model; L-E, linear trend model, exponential correlation model; Q-E, quadratic trend model, exponential correlation model; C-G, zeroth-order trend model, Gaussian correlation model; L-G, linear trend model, Gaussian correlation model; Q-G, quadratic trend model, Gaussian correlation model; C-L, zeroth-order trend model, linear correlation model; L-L, linear trend model, linear correlation model; Q-L, quadratic trend model, linear correlation model; ENS, ensemble of Kriging metamodels.

Table 3: *Normalized actual errors (NRMSE) for individual Kriging metamodels and the ensemble of metamodels for the Branin-Hoo function. The individual Kriging metamodel with the smallest NRMSE is indicated with a bold font*

*N**	*C-E**	*L-E**	*Q-E**	*C-G**	*L-G**	*Q-G**	*C-L**	*L-L**	*Q-L**	*ENS**
10	63.34	78.21	51.61	64.52	68.83	**49.08**	74.73	78.42	53.48	61.49
15	26.75	26.83	25.97	16.51	**15.15**	15.96	69.87	74.08	46.71	16.18
20	21.68	22.75	33.57	15.35	**14.38**	15.11	68.12	68.89	43.62	15.22
25	17.51	17.67	24.74	11.93	11.55	**6.34**	69.65	69.13	41.35	9.16
30	9.46	9.52	11.14	**2.09**	2.20	2.11	69.74	66.65	38.67	2.07
35	5.52	5.47	7.97	0.61	**0.60**	0.64	67.74	64.79	37.56	0.64
40	5.13	5.17	6.58	0.43	0.39	**0.35**	66.92	63.94	36.92	0.36
45	4.94	4.97	5.04	0.35	**0.30**	0.31	67.02	62.85	36.15	0.31
50	4.63	4.66	4.85	0.13	0.12	**0.11**	66.14	62.39	35.90	0.11
55	4.52	4.55	4.73	0.13	0.10	**0.09**	65.96	62.25	35.00	0.09
60	4.54	4.56	4.70	0.079	0.093	**0.077**	65.50	61.85	35.00	0.077

*See footnote below Table 2.

Table 4: *Comparing the accuracy of the ensemble (using NRMSE) to those of the first, second, and third most accurate individual metamodels for different number of training points for the Branin-Hoo function**

*N***	*Better than 1st?*	*Better than 2nd?*	*Better than 3rd?*
10	No	No	Yes
15	No	No	Yes
20	No	No	Yes
25	No	Yes	Yes
30	Yes	Yes	Yes
35	No	No	Yes
40	No	Yes	Yes
45	No	Yes	Yes
50	Yes	Yes	Yes
55	Yes	Yes	Yes
60	Yes	Yes	Yes

*When the error of the ensemble was equal to the error of the individual model, the individual model was taken better than the ensemble.
**Number of training points.

as the number of training points increases, the ensemble becomes as good as the most accurate metamodel.

- Overall, it can be summarized that the best individual Kriging metamodel is usually better than the ensemble. However, which individual metamodel is really the best is unknown in practice. Therefore cross-validation is used, but then the ensemble is always estimated to be the best. The cross-validation gives a good estimate only if 'enough' training points are available.

References

Acar, E. and M. Rais-Rohani (2009) Ensemble of metamodels with optimized weight factors, *Structural and Multidisciplinary Optimization*, **37**, 279–294.

Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*, New York: Oxford University Press, 364–369.

Clarke, S.M., J.H. Griebsch and T.W. Simpson (2005) Analysis of support vector regression for approximation of complex engineering analyses, *Journal of Mechanical Design*, **127**, 1077–1087.

Den Hertog, D., J.P.C. Kleijnen and A.Y.D. Siem (2006) The correct Kriging variance estimated by bootstrapping, *Journal of the Operational Research Society*, **57**, 400–409.

Dixon, L.C.W. and G.P. Szegö (1978) *Towards Global Optimization 2*, Amsterdam, The Netherlands: North Holland.

Forsberg, J. and L. Nilsson (2005) On polynomial response surfaces and kriging for use in structural optimization of crashworthiness, *Structural and Multidisciplinary Optimization*, **29**, 232–243.

Friedman, J.H. (1991) Multivariate adaptive regression splines, *The Annals of Statistics*, **19**, 1–141.

Goel, T., R.T. Haftka, W. Shyy and N.V. Queipo (2007) Ensemble of surrogates, *Structural and Multidisciplinary Optimization*, **33**, 199–216.

Gunn, S.R. (1997) *Support Vector Machines for Classification and Regression, Technical Report*, UK: Image Speech and Intelligent Systems Research Group, University of Southampton.

Hardy, R.L. (1971) Multiquadratic equations of topography and other irregular surfaces, *Journal of Geophysical Research*, **76**, 1905–1915.

Kanazaki, M., K. Tanaka, S. Jeong and K. Yamamoto (2007) Multi-objective aerodynamic exploration of elements' setting for high-lift airfoil using Kriging model, *Journal of Aircraft*, **44**, 858–864.

Kaymaz, I. (2005) Application of Kriging method to structural reliability, *Structural Safety*, **27**, 133–151.

Lebensztajn, L., C.A.R. Marretto, M.C. Costa and J.L. Coulomb (2004) Kriging: a useful tool for electromagnetic device optimization, *IEEE Transactions on Magnetics*, **40**, 1196–1199.

Lee, K-H. and D-H. Kang (2007) Structural optimization of an automotive door using the Kriging interpolation method, *Proceedings of the Institution of Mechanical Engineers – Part D – Journal of Automobile Engineering*, **221**, 1525–1534.

Li, G., S. Azarm, A. Farhang-Mehr and A.R. Diaz (2006) Approximation of multiresponse deterministic engineering simulations: a dependent metamodeling approach, *Structural and Multidisciplinary Optimization*, **31**, 260–269.

Lophaven, S.N., H.B. Nielsen and J. Søndergaard (2002) DACE – A MATLAB Kriging Toolbox, Informatics and Mathematical Modeling, Technical University of Denmark, Lyngby, Denmark.

MacKay, D.J.C. (1998) Introduction to Gaussian processes, in *Neural Networks and Machine Learning*, C.M. Bishop (ed), Vol. **168** of NATO ASI Series, Berlin: Springer, 133–165.

Martin, J.D. and T.W. Simpson (2005) Use of Kriging models to approximate deterministic computer models, *AIAA Journal*, **43**, 853–863.

McLean, P., P. Leger and R. Tinawi (2006) Post-processing of finite element stress fields using dual Kriging based methods for structural analysis of concrete dams, *Finite Elements in Analysis & Design*, **42**, 532–546.

Mourelatos, Z.P., R.C. Kuczera and M. Latcha (2006) An efficient Monte Carlo reliability analysis using global and local metamodels, in *Proceedings of 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 6–8 September, Portsmouth, VA.

Mullur, A.A. and A. Messac (2005) Extended radial basis functions: more flexible and effective metamodeling, *AIAA Journal*, **43**, 1306–1315.

Myers, R.H., D.C. Montgomery and C.M. Anderson-Cook (2009) *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 3rd edn, New York, NY: Wiley.

Sacks, J., W.J. Welch, T.J. Mitchell and H.P. Wynn (1989) Design and analysis of computer experiments, *Statistical Science*, **4**, 409–435.

Santner, T.J., B.J. Williams and W.I. Notz (2003) The design and analysis of computer experiments, New York: Springer-Verlag.

Simpson, T.W., T.M. Mauery, J.J. Korte and F. Mistree (2001a) Kriging models for global approximation in simulation-based multidisciplinary design optimization, *AIAA Journal*, **39**, 2233–2241.

Simpson, T.W., J.D. Peplinski, P.N. Koch and J.K. Allen (2001b) Metamodels for computer based engineering design: survey and recommendations, *Engineering with Computers*, **17**, 129–150.

Smith, M. (1993) *Neural Networks for Statistical Modeling*, New York: Von Nostrand Reinhold.

Stein, M.L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging, Springer Series in Statistics*, New York: Springer-Verlag.

Venkataraman, S. and R.T. Haftka (2004) Structural optimization complexity: what has Moore's law done for us? *Structural and Multidisciplinary Optimization*, **28**, 375–387.

Wang, J.M., D.J. Fleet and A. Hertzmann (2005) Gausssian process dynamical models, in *Proceedings of the 18th Advanced Neural Information Processing Systems Cenference*, December, Vancouver, Canada, 1441–1448.

Wang, X.F., G. Xi and Z.H. Wang (2006) Aerodynamic optimization design of centrifugal compressor's impeller with Kriging model, *Journal of Power and Energy*, **220**, 589–597.

Zerpa, L., N.V. Queipo, S. Pintos and J. Salager (2005) An optimization methodology of alkaline–surfactant–polymer flooding processes using field scale numerical simulation and multiple surrogates, *Journal of Petroleum Science and Engineering*, **47**, 197–208.

The author

Erdem Acar

Erdem Acar received his BS and MS degrees in Aerospace Engineering at the Middle East Technical University in 1999 and 2002, respectively, and his PhD degree in the same field at the University of Florida in 2006. He served as postdoctoral research associate at the Center for Advanced Vehicular Systems of the Mississippi State University from 2006 to 2008. He began his academic career in 2008 at the TOBB University of Economics and Technology of Turkey, where he is currently an Associate Professor working at the Department of Mechanical Engineering. He is actively engaged in teaching and research. His primary areas of teaching include structural analysis and reliability assessment of mechanical systems. He is the author of a book entitled "Aircraft Structural Safety." His research efforts are focused on design optimization of mechanical systems, with strong emphasis on reliability-based design optimization and surrogate-based design optimization. He has published around fifty refereed journal and conference papers in the above areas. He is an associate member of the AIAA Non Deterministic Approaches Technical Committee.